

# Python Machine learning labs

Good morning!

We start at 10:00

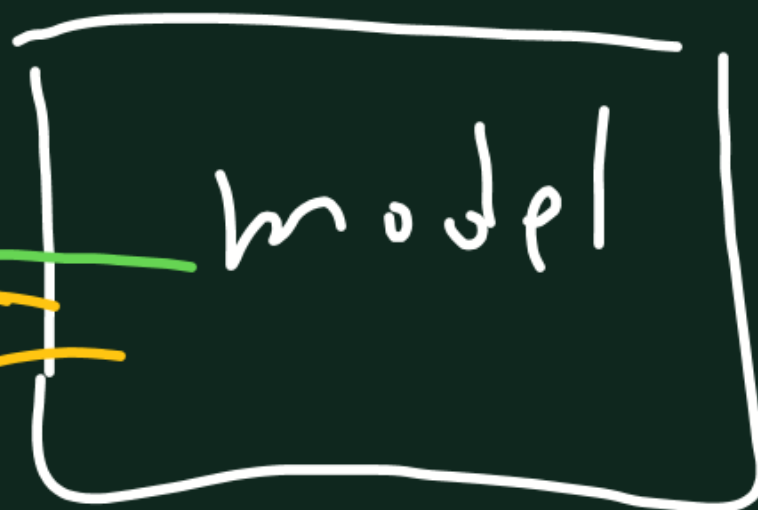


# Exploratory Data Analysis

- Statistical analysis (descriptive statistics)
- Identifying missing values
- Identifying outliers
- Data imbalance
- Plotting

Dataset

# Data Imbalance (1)



CAT

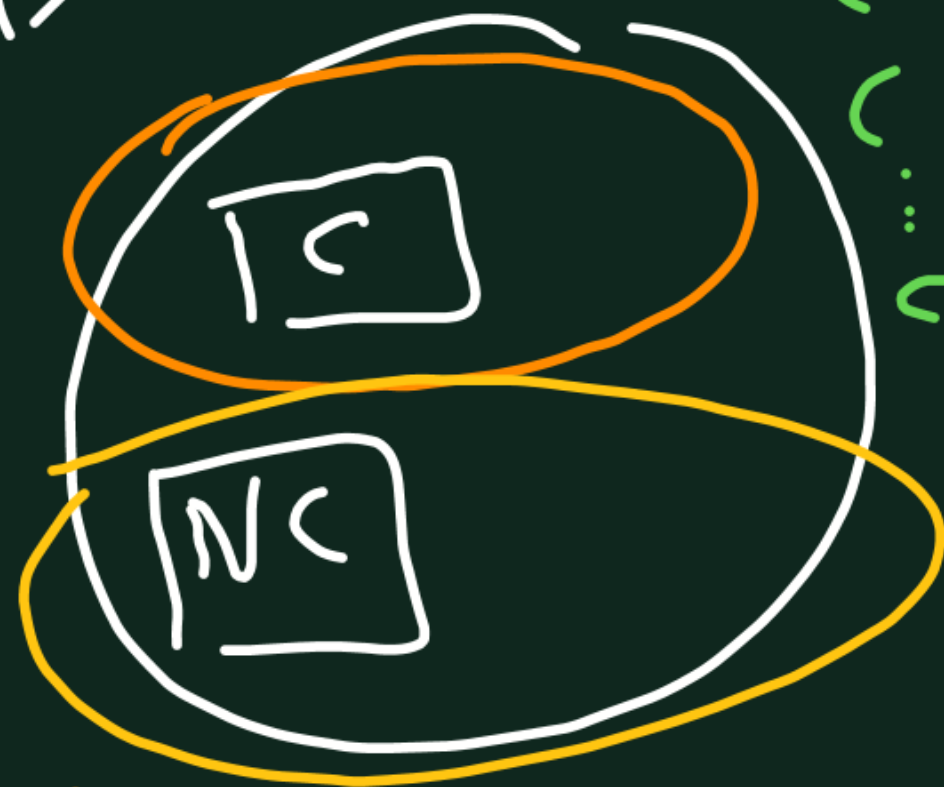
DOG

75% D  
25% C

Accuracy = 75%

# Data Imbalance (2)

Dataset



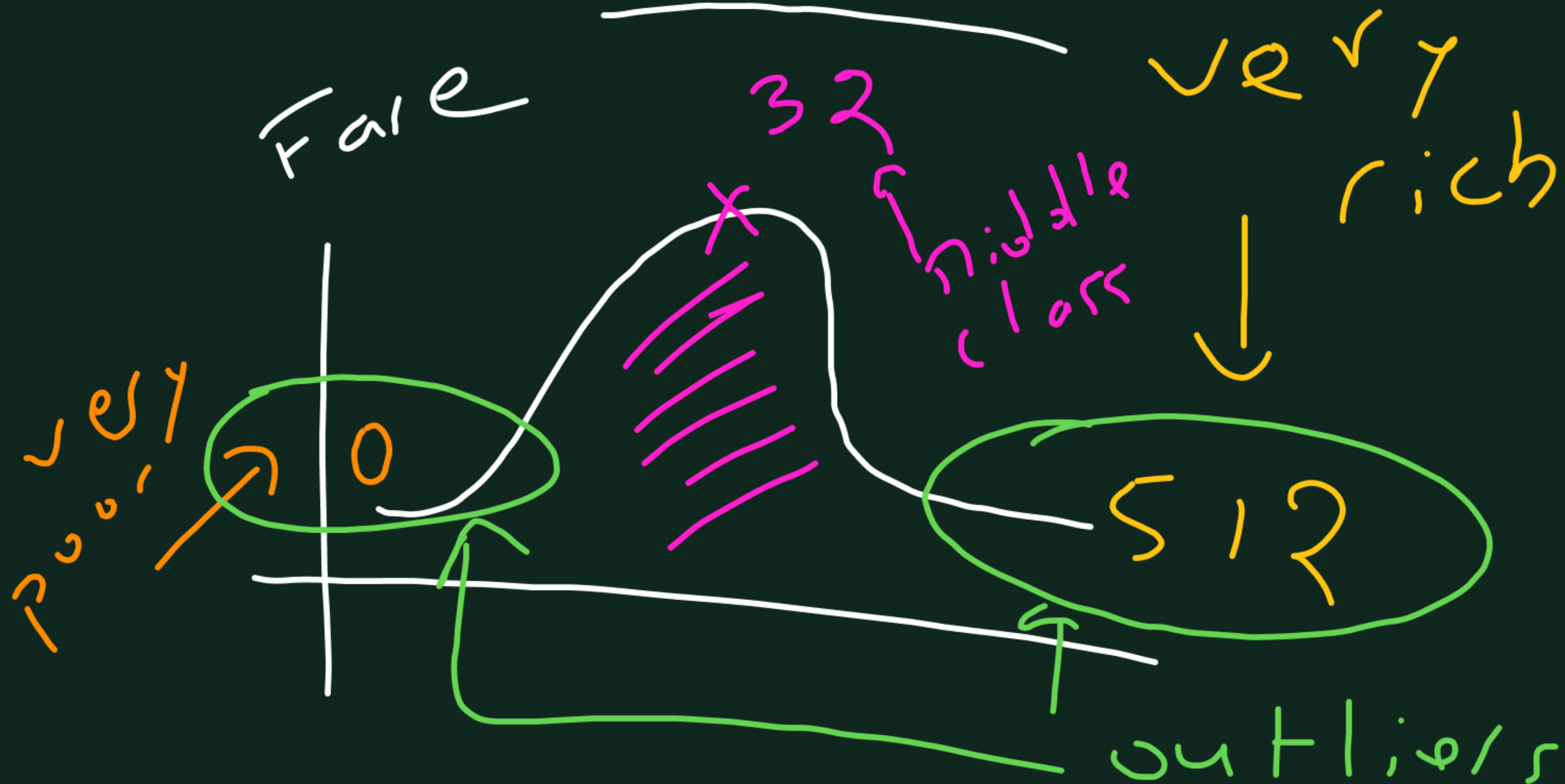
10% → CANCER

90% → No CANCER



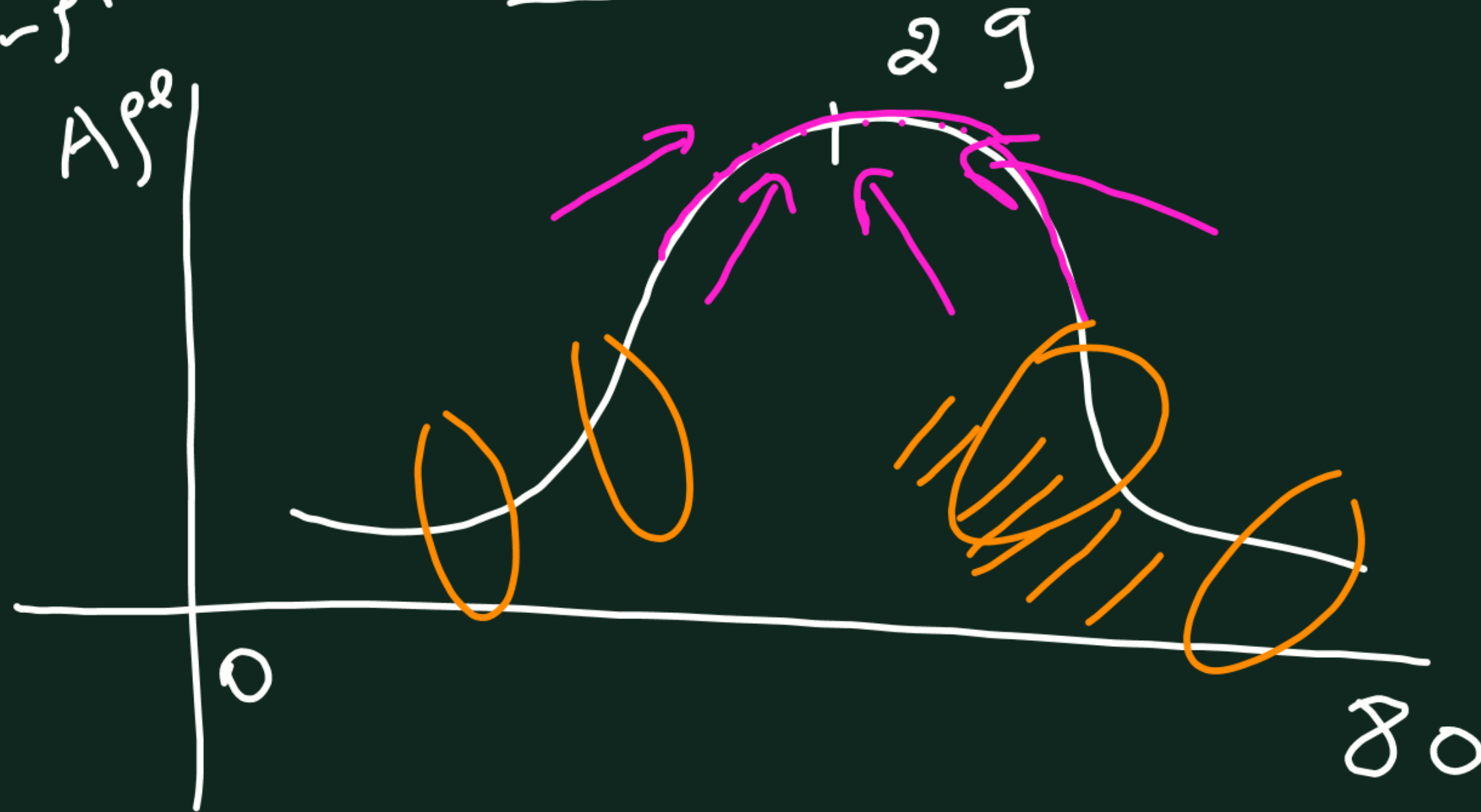
Accuracy = 90%

# Titonic (1)



Feature  
Engineering

Titanic (2)

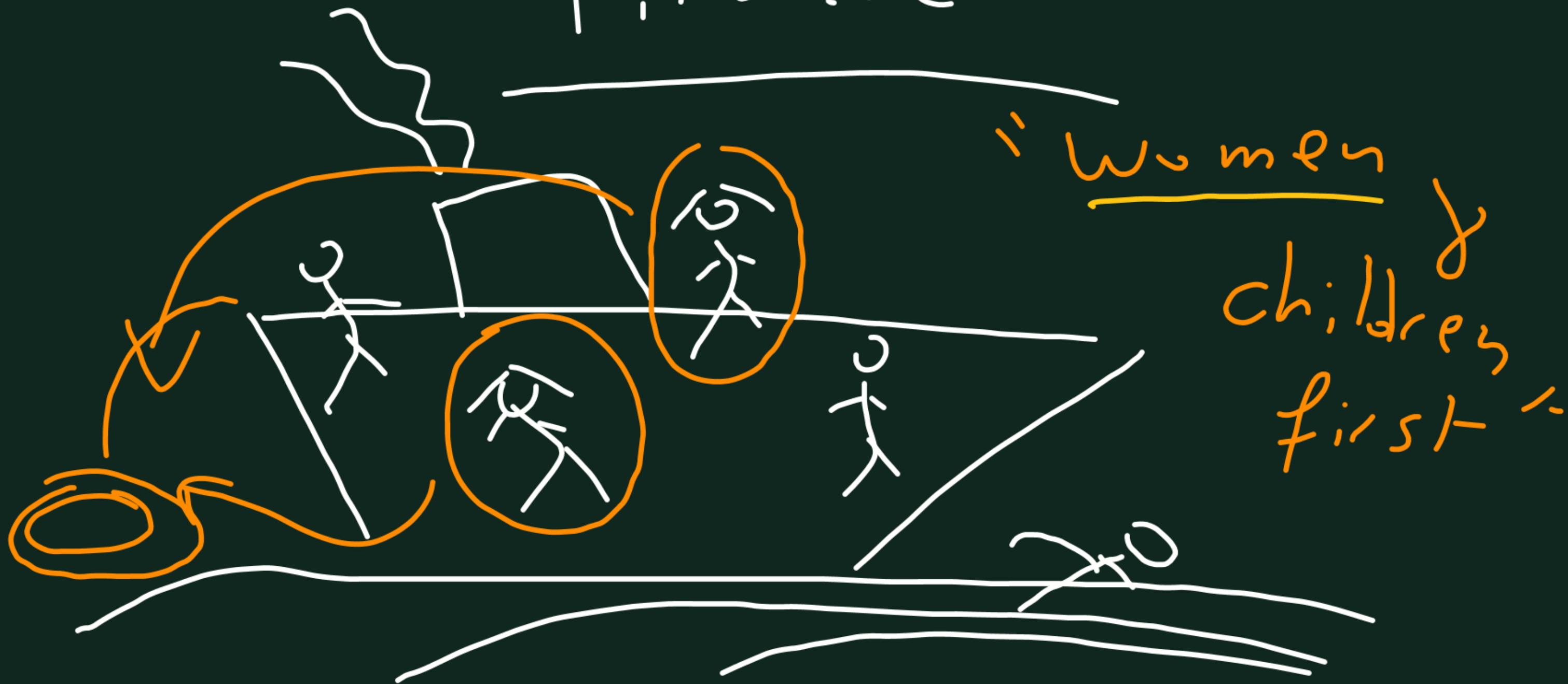


Break

---

Back at 11:30

# Titanic (3)





# Correlation Matrix

Negative

$$-1 < \text{corr} < 0$$



(negative relationship)

No correlation

$$\text{corr} \approx 0$$



(No influence)

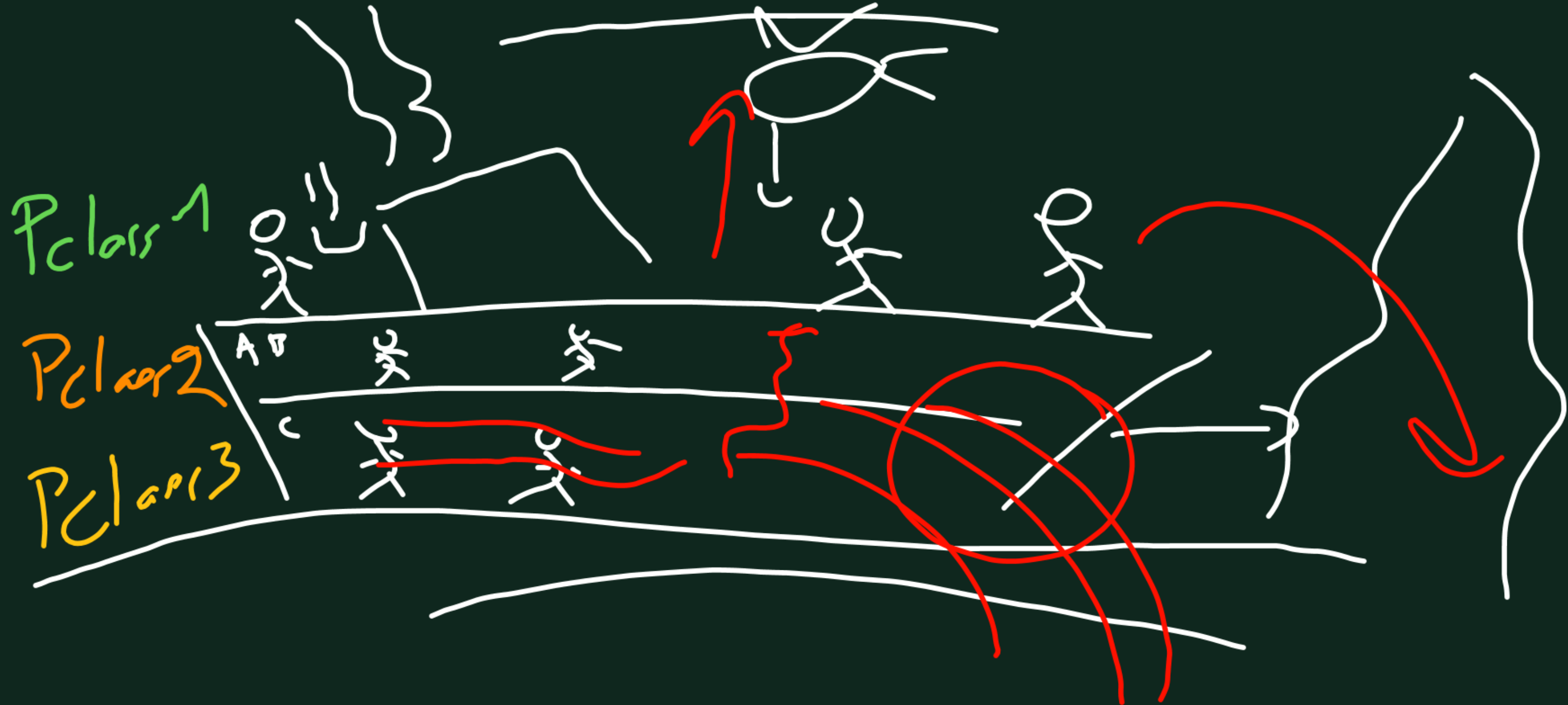
Positive

$$0 < \text{corr} < 1$$



(Positive influence)

# Titanic (4)



Lunch

Back at 13:00

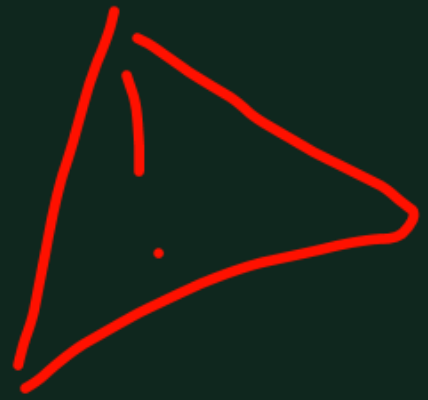
Break

Back at 15:10

Titanic(5)

myname = "Pir/s."

Print(myname[1:])



# Feature Engineering

- Filling (missing values)
- Selection (Dropping columns)
- Transformation (extracting information through functions)